

# The quantum mechanics-based polarizable force field for water simulations

Saber Naserifar, and William A. Goddard

Citation: *J. Chem. Phys.* **149**, 174502 (2018); doi: 10.1063/1.5042658

View online: <https://doi.org/10.1063/1.5042658>

View Table of Contents: <http://aip.scitation.org/toc/jcp/149/17>

Published by the [American Institute of Physics](#)

---

---

**PHYSICS TODAY**

WHITEPAPERS

## ADVANCED LIGHT CURE ADHESIVES

Take a closer look at what these environmentally friendly adhesive systems can do

READ NOW

PRESENTED BY  
**MASTERBOND**  
ADHESIVES | SEALANTS | COATINGS

# The quantum mechanics-based polarizable force field for water simulations

Saber Naserifar and William A. Goddard III<sup>a)</sup>

*Materials and Process Simulation Center, California Institute of Technology,  
Pasadena, California 91125, USA*

(Received 3 June 2018; accepted 13 September 2018; published online 6 November 2018)

We report here a new force field for water based **solely on quantum mechanics (QM)** calculations with **no empirical data**. The QM was at a high level, coupled cluster single double triple, for all orientations and distances for water dimer plus X3LYP density functional theory (DFT) on 19 larger water clusters. In addition, we included charge and polarization based on the polarizable charge equilibration method and nonbond interactions from DFT-D3 calculations on the H<sub>2</sub> and O<sub>2</sub> crystal. This model, denoted as RexPoN, provides quite excellent agreement with experimental (expr) data for the solid and liquid phase of water:  $T_{\text{melt}} = 273.3$  K (expr = 273.15 K) and properties at 298 K:  $\Delta H_{\text{vap}} = 10.36$  kcal/mol (expr = 10.52), density = 0.9965 gr/cm<sup>3</sup> (expr = 0.9965), entropy = 68.4 (J/mol)/K (expr = 69.9), dielectric constant = 76.1 (expr = 78.4), and  $\ln D_s$  (self-diffusion coef) = -10.08 (expr = -11.24). Such an accurate force field for water will, we believe, be useful for full solvent calculations of electrocatalysis, where we can restrict QM water to just the first one or two layers involving reactions, using RexPoN to provide the polarization for a more distant solvent. Also, RexPoN may provide a better description of the solvent for proteins, DNA, polymers, and inorganic systems for applications to biomolecular, pharma, electrocatalysis (fuel cells and water splitting), and batteries where interaction with explicit water molecules plays a significant role. *Published by AIP Publishing.* <https://doi.org/10.1063/1.5042658>

## I. INTRODUCTION

Because of its central role in life and sustenance, the properties of water have been widely studied. Although chemically simple, it has been hard to model<sup>1,2</sup> with many anomalies such as the critical point of supercooled water below 219 K,<sup>3</sup> the maximum density at 4 °C, the high dielectric constant and boiling point, the non-monotonic isothermal compressibility and specific heat,<sup>4</sup> and the fifteen known crystal structures for ice.<sup>5,6</sup> Because of the lack of long range order, molecular dynamics (MD) simulations have played the major role in elucidating these properties. These MD studies are nearly all based on empirical force fields (FFs) because quantum mechanics (QM) simulations are limited to 100s of atoms for 10s of picoseconds (ps). Starting with the first model of Bernal and Fowler in 1933 and the first computer simulation of water by Barker and Watts in 1969, hundreds of water models have been developed and applied to the properties of water<sup>1,2</sup> but there remain many puzzles.

The most popular empirical FFs<sup>1,2</sup> are the TIPnP (n = 3, 4, 5)<sup>7-9</sup> and simple point-charge (SPC)<sup>10</sup> families, which use rigid molecular structures. The parameters in these models are fitted to reproduce one or more macroscopic properties of water but none reproduce all important properties of the solid and liquid phase.

Flexible (e.g., TIP4PF<sup>11</sup> and SPC/Fw<sup>12</sup>) and fluctuating charge (e.g., TIP4PQ<sup>77</sup> and SPC/FQ<sup>13</sup>) models are used less

often because of increased costs with no obvious increase in accuracy.

Polarizable FFs [e.g., Anisotropic Site Potentials (ASP);<sup>14</sup> Symmetry-adapted Perturbation Theory (SAPT);<sup>15</sup> and Thole-type Model (TTM) families<sup>16</sup>] and Drude oscillator methods (e.g., SWM4-DP<sup>17</sup>) have been used a little, but some exhibit a “polarization catastrophe” at short distances.<sup>18</sup>

Some water models have been developed from fitting many body potential energy terms to millions of high-quality energy datasets [e.g., Huang, Braams, and Bowman (HBB);<sup>19</sup> CC-pol;<sup>20</sup> and MB-pol<sup>21</sup>]. These models are not easily transferable to other systems since for any new set of interactions (new species), this tedious fitting procedure must be repeated.<sup>22</sup>

*Ab initio* molecular dynamics (AIMD) methods have been improved over the last two decades to perform at least close to empirical models.<sup>23</sup> Various computational techniques have been implemented to improve the qualities of water models but not yet successful. For example, machine learning methods have been used to improve the quality of explicit polarization and Coulombic terms in the water models.<sup>24</sup> More discussion of different types of water models is provided elsewhere.<sup>1,2</sup>

We decided to develop a FF based fully on the best available QM, with the hope that it would correctly describe the standard properties of water including the melting point ( $T_m$ ), density ( $\rho$ ), heat of vaporization ( $\Delta H_{\text{vap}}$ ), entropy ( $S^0$ ), dielectric constant ( $\epsilon$ ), self-diffusion constant ( $D_s$ ), oxygen-oxygen radial distribution function ( $R_{\text{OO}}$ ), and which might additionally explain the anomalies of water. Our procedure is as follows:

<sup>a)</sup>Author to whom correspondence should be addressed: wag@caltech.edu

- First, establish the **long range interactions**: the van der Waals nonbond (NB) interactions (London dispersion plus Pauli repulsion) from QM.
- Describe the electrostatic (polarization and charge) interactions based on our recently developed polarizable charge equilibration (PQEq) method that has been shown to reproduce the QM polarization energies due to bringing point dipole up to various small molecules.
- Use Coupled Cluster Single Double Triple, CCSD(T), QM on a water monomer<sup>25</sup> (including OH bond dissociation and HOH angle terms) to determine the **bond and angle valence terms**.
- Use Coupled Cluster Single Double Triple, CCSD(T), QM on a water dimer<sup>25</sup> to determine the **hydrogen bond (HB) corrections**.

The final FF is called RexPoN to indicate that it includes charge and polarization (Po) and Nonbond (N) while allowing bond breaking (Rex). RexPoN is aimed at doing both reactive and nonreactive systems. Thus, we use the full RexPoN formalism to define the valence bond and angle terms at the same level as would be used in reactions.

However, in the validations reported here, we use a rigid water model where OH bonds and HOH angles are kept fixed at the values from CCSDT throughout the simulations (see below). There are questions concerning whether the rigid water model describes nuclear quantum effects (NQE) including zero-point energy (ZPE) and tunneling. The hydrogen bonding between water molecules can be affected by ZPE effects for OH bond stretch and HOH bending vibrational modes.<sup>26</sup> However, it has been shown that previous estimates of NQE effects were greatly overestimated because of the use of harmonic intramolecular interactions, ignoring anharmonicity.<sup>27</sup> In addition, proton delocalization and tunneling effects are expected to be important only at very low temperatures and high pressures which are not the conditions of simulations here.<sup>26</sup> Moreover, RexPoN includes the correct anharmonicity at the level of CCSDT since the OH bond and HOH angle curves were fitted to this level. Therefore, flexible RexPoN will likely do well on NQE effects in water.

The rest of this paper is organized as below. Section II describes the model and the physical motivation behind each potential energy term. Section II also discusses the training sets and optimization process of the FF. Section III provides the results of the simulations for several ice Ih and water properties and compares the results with experimental data. The conclusions are given in Sec. IV.

## II. THE MODEL

The total energy of the system is expressed as

$$E_{total} = E_{nonbond} + E_{valence}, \quad (1)$$

where  $E_{nonbond}$  is the nonbond (NB) energy and  $E_{valence}$  is the bonded energy terms.  $E_{nonbond}$  and  $E_{valence}$  are defined by

$$E_{nonbond} = E_{elect} + E_{vdW} + E_{HB}, \quad (2)$$

$$E_{valence} = E_{bond} + E_{angle}, \quad (3)$$

where  $E_{elect}$  is the electrostatic,  $E_{vdW}$  is the van der Waals (vdW), and  $E_{HB}$  is the hydrogen bond (HB) energy.  $E_{valence}$  includes the covalent OH bond ( $E_{bond}$ ) and the HOH angle ( $E_{angle}$ ) energy terms.

### A. Electrostatic energy ( $E_{elect}$ )

We describe the electrostatic energy of the system using our polarizable charge equilibration (PQEq) model.<sup>28,29</sup> In addition to atomic charges, we use PQEq to include the polarization arising from interatomic interactions and external electric fields. Each atom is considered to have a core including all the mass of the atom and an electron shell that is massless. The core and shell are described by 1s Gaussian shaped electron densities ( $\rho$ ) with a finite distribution of charge on each atom rather than fixed point charges used by many other FFs (see Fig. 1). The core has a variable charge ( $q_i$ ) and a fixed charge (+1) part. The shell only has a fixed charge of  $-1$ . The dynamic position of the shell is found by balancing the external forces due to other atoms and applied fields with the spring force constant ( $K_s$ ) that connects the shell and core of the atom. The electrostatic interaction energy between the electron densities ( $\rho$ ) of atoms  $i$  and  $j$  is given by

$$C_{ik,jl}(r) = \frac{1}{r} \text{erf} \left( \sqrt{\frac{\alpha_{ik}\alpha_{jl}}{\alpha_{ik} + \alpha_{jl}}} r \right), \quad (4)$$

where  $k$  and  $l$  denote the core (c) and shell (s), respectively. The width of the Gaussian distribution ( $\alpha_{ik}$  and  $\alpha_{ik}$ ) is equal to  $0.2341/R^2$ , where  $R$  is the core ( $R_c$ ) or shell ( $R_s$ ) radius. The electrostatic energy ( $E_{elect}$ ) is defined by

$$E_{elect}(\{r_{ic}, r_{is}, q_i\}) = \sum_i^N \left[ \chi_i^0 q_i + \frac{1}{2} J_{ii}^0 q_i^2 + \frac{1}{2} K_s r_{ic, is}^2 \right] + \sum_{ik>jl} T(r_{ik,jl}) C_{ik,jl}(r_{ik,jl}) q_{ik} q_{jl}, \quad (5)$$

where  $\chi_i^0$  is the Mulliken electronegativity,  $\chi_i^0 = (\text{IP} + \text{EA})/2$ ,  $J_{ii}^0$  is the idempotential,  $J_{ii}^0 = \text{IP} - \text{EA}$ ,  $r_{ik,jl}$  is the interatomic distance, IP = atomic ionization potential, and EA = atomic electron affinity. Here  $T(r)$  is the 7th order taper function [see Eq. (29)]. The second sum computes the electrostatic energy

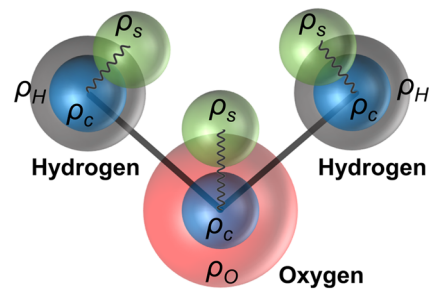


FIG. 1. The PQEq model for a water molecule. Each atom has a core and shell. The core has a variable ( $\rho_H$  and  $\rho_O$ ) and a fixed charge distribution ( $\rho_c$ ) both positioned at the center of the atom. The shell has a fixed shell charge distribution ( $\rho_s$ ) with negative charge connected by a harmonic spring to  $\rho_c$ . The  $\rho_c$  and  $\rho_s$  have the overall charges of +1 and  $-1$ , respectively. The figure shows the positions of shells for a single isolated water molecule. The position of each shell is found by balancing the external forces due to other atoms with the spring force that connects the shell and core of the atom.

TABLE I. PQEq parameters for hydrogen and oxygen atoms in a water molecule.

Atom	$\chi_i^0$ (eV)	$J_{ii}^0$ (eV)	$R_c = R_s$ (Å)	$K_s$ [(kcal/mol)/Å <sup>2</sup> ]
H	4.5280	17.9841	0.4452	2037.2006
O	8.7410	13.3640	0.8028	814.0445

between the core and shells of all atoms in the system. The charge on the core ( $q_i$ ) is updated every time step via efficient schemes. The massless shell relaxes instantaneously to align itself along the external force vector with no inertial delay. The complete description of the PQEq model and each term in the energy equation is provided in previous publications.<sup>28,29</sup> We validated the accuracy of PQEq by comparing with the polarization in QM.<sup>28</sup> We reoptimized the PQEq parameters of water for two reasons.

First, to ensure that the charge distribution on oxygen and hydrogen atoms is in reasonable ranges when two water molecules get very close to each other. This is important for describing the equation of state (EOS) of water at high pressures. Thus, we computed the QM [electrostatic potential (ESP) and Mulliken] charge distributions on the water dimer over a range of short distances and adjusted the PQEq parameters such that they produce reasonable charges (see Fig. S1 of the [supplementary material](#)).

Second, during the optimization of HB attractive and repulsive terms (see below), we optimize the PQEq parameters such that the electrostatics energy provides the most accurate description of the water dimer energy at long distances. The PQEq parameters of H and O are provided in Table I.

## B. van der Waals energy ( $E_{vdW}$ )

To determine the vdW nonbond interactions, we calculated the equation of states (EOS) of solid hydrogen and solid oxygen based on the density functional theory (DFT).<sup>30,31</sup> For solid hydrogen, we use the Perdew-Burke-Ernzerhof (PBE)<sup>32</sup> flavor of DFT, while for a solid oxygen crystal, we use Becke three-parameter Lee-Yang-Parr (B3LYP)<sup>33,34</sup> including the hybrid terms important for describing the O<sub>2</sub> triplet state<sup>35,36</sup> (see below).

### 1. Solid hydrogen

X-ray diffraction shows that the H<sub>2</sub> crystal has a hexagonal closed packed structure with the  $P6_3/mmc$  space group and a  $c/a$  ratio close to the ideal value of 1.633.<sup>37</sup> However, since hydrogen has no core electrons, X-ray experiments cannot determine the exact position of the H atom coordinates. Therefore, extensive theoretical studies have been carried out to determine the H<sub>2</sub> molecule orientation. It was found that the  $Pca2_1$  space group (see Fig. S2 of the [supplementary material](#)) containing a hexagonal close packed structure with four molecules per unit cell is the most stable structure at high pressure ranges (~110–150 GPa).<sup>38,39</sup> Therefore, we selected a  $Pca2_1$  space group and used the PBE-D3 to compute energy versus volume with single point energy calculations at

various volumes. This leads to the PBE-D3 energy curve shown in Fig. 2(a) for densities starting with the volume of 17.440 Å<sup>3</sup> ( $\rho = 2.626$  cm<sup>3</sup>/mol) corresponding to the pressure of 120 GPa at 300 K and passing through the volume of 44.082 Å<sup>3</sup> (6.627 cm<sup>3</sup>/mol) corresponding to the pressure of 10 GPa at 300 K. We also scale the simulation cell to large volumes as high as 424.572 Å<sup>3</sup> (63.920 cm<sup>3</sup>/mol) to capture the long-range interactions. The HH bond distance was fixed at the equilibrium value of 0.74 Å for all volumes to avoid any contribution to the total PBE-D3 curve from changes in the HH bond length. All calculations in this section were performed using the Vienna *Ab initio* simulation package (VASP) (version 5.3.5).<sup>40</sup> The generalized gradient approximation (GGA) was used for the exchange–correlation energy<sup>32</sup> with a plane wave energy cutoff of 600 eV. We used the convergence criteria of 10<sup>−6</sup> eV for energy and 10<sup>−4</sup> eV Å<sup>−1</sup> for force. The k-space sampling was the gamma point centered with full space group symmetry.

### 2. Solid oxygen

For solid oxygen, we selected the stable  $\alpha$ -phase crystal structure, which according to neutron-diffraction and X-ray measurements has a monoclinic base-center structure with a  $C2/m$  space group<sup>35,36</sup> (see Fig. S2 of the [supplementary material](#)). For DFT, we used the B3LYP hybrid method. This

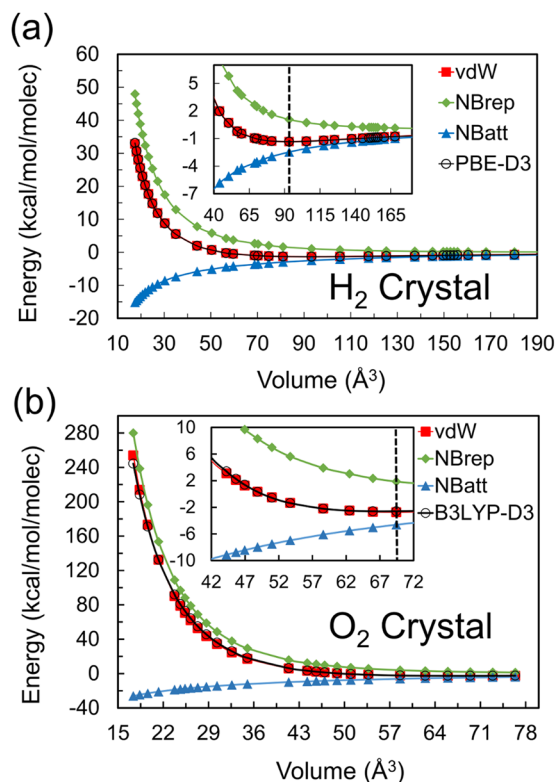


FIG. 2. The change of vdW energies with volume for (a) solid hydrogen and (b) solid oxygen computed by DFT-D3 (open circles) and RexPoN (red squares). The vdW curves are the sum of the nonbond repulsion (NBrep) and nonbond attraction (NBatt). The dotted vertical line in the insets shows the volume for PBE-D3 for the minimum energy. The DFT-D3 energy in (a) is computed by PBE-D3, and the volumes of 17.440 and 44.082 Å<sup>3</sup> are related to the experimental pressures of 120 and 10 GPa at 300 K, respectively. The DFT-D3 energy in (b) is computed by B3LYP-D3, and the volumes of 17.369 and 41.376 Å<sup>3</sup> are related to the experimental pressures of 300 and 10 GPa at 300 K, respectively.

is because  $O_2$  has a  $^3\Sigma_g^-$  triplet ground state that is poorly described by PBE. The ground state of  $O_2$  crystal has the two  $O_2$  molecules coupled antiferromagnetically to obtain a singlet spin state. Since the DFT formalism uses a single Slater determinant, it cannot properly describe the single spin state, often leading to severe convergence problems. Instead, we describe the  $O_2$  crystal assuming a high spin state quintet state ( $S = 2$ , with 4 extra up spin orbitals per cell,  $M_S = 2$ ) which is correctly described with a single Slater determinant using a spin unrestricted DFT formalism. Our computed heat of vaporization of solid oxygen using B3LYP-D3 and high spin  $O_2$  is  $-2.3$  kcal/mol after correcting for zero-point energy, which is very close to the experimental value of  $-2.1$  kcal/mol.<sup>36</sup> The EOS of solid oxygen was then obtained by performing single point energy calculation at different volumes. This leads to the B3LYP-D3 energy curve shown in Fig. 2(b). The smallest volume of  $17.369 \text{ \AA}^3$  ( $\rho = 5.230 \text{ cm}^3/\text{mol}$ ) corresponds to a pressure of 300 GPa at 300 K,<sup>41,42</sup> the volume of  $41.376 \text{ \AA}^3$  ( $12.459 \text{ cm}^3/\text{mol}$ ) corresponds to a pressure of 10 GPa at 10 K,<sup>43</sup> and the volume  $69.439 \text{ \AA}^3$  ( $20.909 \text{ cm}^3/\text{mol}$ ) corresponds to a pressure of 0 GPa at 10 K.<sup>43</sup> The OO bond distance was fixed at the equilibrium value of  $1.207 \text{ \AA}$ . All calculations in this section were performed using the CRYSTAL package (version 14)<sup>44</sup> using the modified Gaussian basis set, m-6-311G(2d) for oxygen.<sup>45</sup> We used the convergence criteria of  $10^{-7}$  a.u. for energy. The k-space sampling was the gamma point centered with full space group symmetry.

We used the CRYSTAL package for solid oxygen because the VASP package is over 1000 times slower than CRYSTAL for B3LYP calculations. However, we used VASP for PBE-D3 calculations of solid hydrogen because it is far easier to use than CRYSTAL and is only a factor of 3 slower.

The hydrogen and oxygen DFT-D3 energy curves were then fitted to obtain the first principles based vdW potential energy curves for H and O. This vdW energy is the sum of a monotonic attractive nonbond term (NBatt) and a monotonic repulsive non-bond term (NBrep). NBatt has the form of  $-C_6/r^6$  which means that at sufficiently long  $r$  the wavefunction on separate atoms does not overlap. The NBatt must remain finite as the distances become sufficiently small. Therefore, we define NBatt energy ( $E_{NBatt}$ ) as

$$E_{NBatt}(r) = - \sum_{i<j} \frac{C_{6ij}}{r_{ij}^6 + R_{vdWij}^6}, \quad (6)$$

which is similar to low gradient formulation<sup>46,47</sup> that was also used in ReaxFF-Ig.<sup>48</sup> Here,  $r$  is the interatomic distance,  $C_6$  is the dispersion energy correction parameter, and  $R_{vdW}$  is the equilibrium vdW distance between atoms  $i$  and  $j$ . The NBrep energy ( $E_{NBrep}$ ) is purely repulsive (positive)

TABLE II.  $E_{NBatt}$  parameters of the vdW energy term of the water model.

Atom $i$	Atom $j$	$R_e$ (Å)	$C_6$ [(kcal Å <sup>6</sup> )/mol]
H	H	3.2692	239.4803
O	O	2.9938	634.7066
O	H	3.1285	389.8714

and monotonic which we expand using a group of exponential functions as

$$E_{NBrep}(r) = \sum_{i<j} A_{ij} \exp(\alpha_{ij} r_{ij} + \beta_{ij}) \exp(\gamma_{ij} r_{ij}^{n_{ij}} + \eta_{ij} r_{ij} + \delta_{ij}), \quad (7)$$

where  $A$ ,  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $n$ ,  $\eta$ , and  $\delta$  are the parameters.

In developing RexPoN, we want to use generic approaches likely to have the correct systematics as a function of the rows and columns of the periodic table. Thus we used as initial values for  $R_{vdW}$  and  $C_6$  the values from the universal force field,<sup>49</sup> which has parameters up to Lw ( $Z = 103$ ). We expect that the  $R_{vdW}$  should be reasonably accurate; however, the  $C_6$  was based on calculated atomic polarizabilities and needs to be adjusted. Next, we define the initial parameters of  $E_{NBrep}$  knowing that  $E_{NBrep}(r) = E_{vdW}(r) - E_{NBatt}(r)$ . Then, the parameters of  $E_{NBatt}$  and  $E_{NBrep}$  are optimized to provide the best fit to the  $E_{vdW}$  curve.

To describe HO vdW interaction, we used the geometric mean average of the two-body vdW potential energy curves of H and O as in the following equations:

$$E_{NBrep}(r_{OH}) = \sqrt{E_{NBrep}(r_{HH})E_{NBrep}(r_{OO})}, \quad (8)$$

$$R_{vdW,OH} = \sqrt{R_{vdW,HH}R_{vdW,OO}}, \quad (9)$$

$$C_{6,OH} = \sqrt{C_{6,HH}C_{6,OO}}. \quad (10)$$

The final  $E_{NBatt}$  and  $E_{NBrep}$  parameters are given in Tables II and III, respectively.

### 3. Validation of the method

We tested the accuracy of DFT-D3 by comparing to the experimental pressure-volume ( $P$ - $V$ ) EOS of solid hydrogen. We used experimental measurements on solid hydrogen isotopes ( $H_2$  and  $D_2$ ) within the pressure range of 10–120 GPa at 300 K for these comparisons<sup>37</sup> (Fig. 3). We first minimized the atomic positions related to this pressure range (i.e., from  $17.440$  to  $44.082 \text{ \AA}^3$ ) using PBE-D3 to obtain the total energy at 0 K. Then, we corrected this energy for the zero point energy (ZPE) contributions and also free energy (thermal effects) at 300 K ( $G = H - TS$ ). The total PBE-D3 energy before

TABLE III.  $E_{NBrep}$  parameters of the vdW energy term of the water model.

Atom $i$	Atom $j$	$A$ (kcal/mol)	$\alpha$	$\beta$	$\gamma$	$n$	$\eta$	$\delta$
H	H	0.0212	-2.4002	10.5460	-0.8113	0.0014	-0.2607	-0.5640
O	O	0.0272	-3.6468	14.5278	-0.3235	0.0305	0.0331	-0.5688
O	H	0.0906	-3.1455	10.0823	-0.0007	-1.0000	0.0000	0.0000



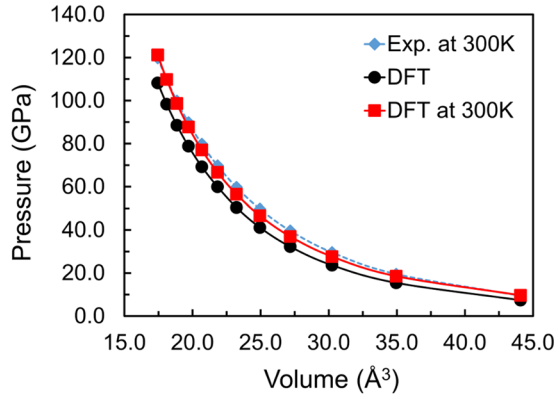


FIG. 3. The computed and experimental pressure-volume equation of state of the hydrogen. The DFT curve (black) has been corrected for zero-point energy and thermal effects at 300 K (red) to compare with the experimental curve<sup>37</sup> (blue) at 300 K.

and after correction at 300 K is shown in Fig. S4 of the [supplementary material](#). Then, the total energy curve at 300 K was fitted to third-order Birch-Murnaghan isothermal EOS,<sup>50</sup>  $E_{BM}(V)$ . The  $P$ - $V$  EOS was then obtained by

$$P_{BM}(V) = -\left(\frac{\partial E_{BM}}{\partial V}\right)_S. \quad (11)$$

As shown in Fig. 3, our PBE-D3 predicted EOS is in good agreement with experiment, indicating that DFT-D3 is sufficiently accurate to describe the vdW interactions of dense solid systems.

### C. Covalent bond energy ( $E_{bond}$ )

#### 1. The bond order (BO) expression

To describe the covalent O–H bond, we used the bond order (BO) concept to capture the dependence of the bonding on the overlap (actually exchange kinetic energy) between orbitals on the bonding atoms. Here we use a sigmoid function as in ReaxFF<sup>51</sup> to define the bond distance to the bond order relation

$$BO_{ij} = \frac{1.10}{1.0 + \exp\left[-p_{bo1}\left(\frac{r_{ij}-p_{bo2}}{R_e-p_{bo2}}\right)\right]}, \quad (12)$$

where  $p_{bo1}$  and  $p_{bo2}$  are the parameters,  $R_e$  is the equilibrium bond distance, and  $r$  is the bond distance between atom  $i$  and atom  $j$ . We choose the  $BO = 1.0$  at  $R_e$  which results in  $p_{bo1} = 2.3026$ . This allows the formal bond order to approach 1.1 for very small values of  $r$ . Therefore, the only free parameter that needs to be determined is  $p_{bo2}$ . To make an initial good guess for  $p_{bo2}$ , we choose the  $BO = 0.5$  at the inflection point distance ( $R_{H/2}$ ) of the bond energy curve. The QM bond energy curves (see below) are used to obtain  $p_{bo2}$  parameters for  $H_2$ ,  $O_2$ , and OH (in  $H_2O$  molecule) cases. The final bond order curves are shown in Fig. S5 of the [supplementary material](#), and the BO parameters are summarized in Table IV.

We define the bond energy ( $E_{bond}$ ) to match the bond dissociation curves in  $H_2$ ,  $O_2$ , and OH (in  $H_2O$ ) molecules for the highest quality QM calculations. Thus, given  $E_{vdw}$  and  $E_{elect}$ , we define  $E_{Q2}$  and  $E_{bond}$  as

TABLE IV. Bond order parameters for HH, OO, and OH bonds.

Bond type	$R_e$ (Å)	$p_{bo1}$	$p_{bo2}$
H–H	0.7450	2.3026	1.3532
O–O	1.2070	2.3026	1.6150
O–H	0.9587	2.3026	1.4428

$$E_{Q2}(r_{ij}) = E_{QM}(r_{ij}) - \left[ \sum_{i,j \neq N}^N E_{vdW}(r_{kl}) + \sum^N E_{elect}(r_{kl}) \right], \quad (13)$$

$$E_{bond}(r_{ij}) = E_{Q2}(r_{ij}) - E_{vdW}(r_{ij}), \quad (14)$$

where  $E_{QM}$  is the total QM energy due to the stretching bond between atom  $i$  and  $j$ .  $k$  and  $l$  are the atom indices ( $k, l \neq i, j$ ).  $E_{Q2}$  is the net two-body QM energy between atom  $i$  and  $j$ , which is obtained by subtracting from  $E_{QM}$  the  $E_{vdW}$  between all atom pairs except for  $i$  and  $j$  and also subtracting the total  $E_{elect}$  of the system [Eq. (5)]. Therefore, based on  $E_{vdW}$  and  $E_{electro}$  which are already established, the  $E_{bond}$  is defined automatically to match the QM bonding energy curves.

We used the exact *ab initio* singlet energy data for the HH bond scan in a  $H_2$  molecule<sup>52</sup> and high quality *ab initio* and spectroscopic energy data for the OO bond scan in an  $O_2$  molecule.<sup>53,54</sup> For diatomic  $H_2$  and  $O_2$  molecules, the  $E_{vdW}(r_{kl})$  and  $E_{electro}(r_{kl})$  sums in Eq. (13) are zero and therefore  $E_{Q2}(r_{ij}) = E_{QM}(r_{ij})$ . For the OH bond in the  $H_2O$  molecule, we used the accurate water monomer potential energy surfaces (Bowman PES) obtained from CCSD(T) calculations.<sup>25</sup> Here, we started from the equilibrium water structure ( $OH = 0.9572$  Å and  $\angle HOH = 104.52^\circ$ ) and stretched one of the OH bonds to infinity while the angle and other bond are fixed. Then, the  $E_{QM}$  was computed at each distance using Bowman PES. Here, the HH  $E_{vdW}$  and total  $E_{electro}$  [Eq. (13)] should be subtracted from  $E_{QM}$  to get the net OH two-body curve ( $E_{Q2}$ ). The  $E_{Q2}$  curve is fitted to an extended-Rydberg type function of the form

$$E_{Ryd}(r_{ij}) = -D_0 \exp[-b_0 a_0(r - R_e)] \{c_0 + b_0 a_0(r - R_e) \times [c_0 + a_0(r - R_e)(c_0 + a_0(r - R_e))]\}, \quad (15)$$

where  $D_0$ ,  $a_0$ ,  $b_0$ , and  $c_0$  are the parameters to be determined. Figure 4 shows the energy components and the fit to the  $E_{Q2}$  curve for  $H_2$ ,  $O_2$ , and OH bond dissociation. The final parameters of Eq. (15) are given in Table V.

#### D. Angle energy ( $E_{angle}$ )

To define the angle term, we fit a simple cosine angle function to the Bowman PES for the water monomer,<sup>25</sup>

$$E_{angle}(r_{ij}, r_{jk}, \theta_{ijk}) = \frac{2C}{\sin^2(\theta_0)} F_1(BO_{ij}) F_2(BO_{jk}) \times [\cos(\theta_{ijk}) - \cos(\theta_0)]^2, \quad (16)$$

$$F_1(BO_{ij}) = 1 - \exp(-a BO_{ij}^b), \quad (17)$$

$$F_2(BO_{jk}) = 1 - \exp(-a BO_{jk}^b), \quad (18)$$

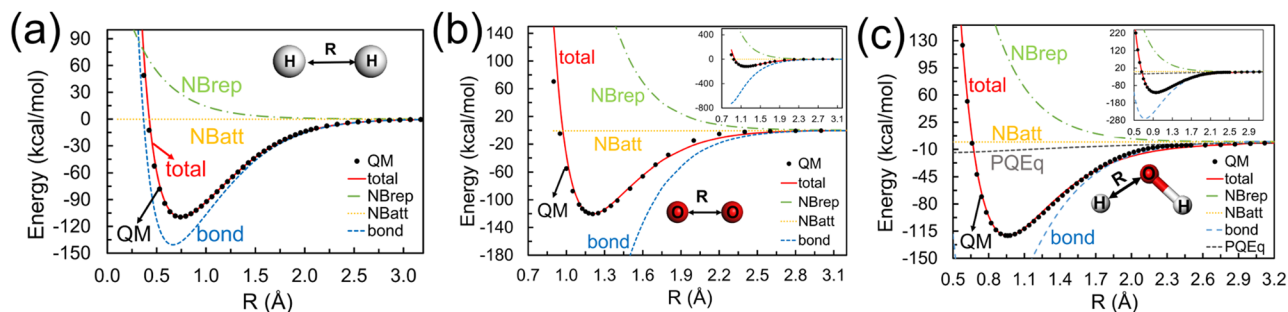


FIG. 4. The QM (black circles) and RexPoN (red) total energies for bond dissociations of (a) HH, (b) OO, and (c) OH. The energy components of RexPoN are also shown. Insets in (b) and (c) show the curves for a broader range such that the complete range of the bond energy curve (blue) is shown.

where  $C$ ,  $\theta_0$ ,  $a$ , and  $b$  are the adjustable parameters.  $\theta_{ijk}$  is the angle between atoms  $i$ ,  $j$ , and  $k$ .  $F_1$  and  $F_2$  are the scaling functions. The scaling functions go to zero as OH gets very large and become equal to one at the equilibrium bond length ( $r_{ij} = R_e$  or  $r_{jk} = R_e$ ). The QM energy as a function of  $\theta_{\text{HOH}}$  with both bonds fixed at  $R_e = 0.9572$  Å (i.e.,  $F_1 = F_2 = 1$ ) is shown in Fig. 5. The angle term together with  $E_{\text{vdw}}$  and  $E_{\text{electro}}$  is fitted to the QM curve. The parameters of angle energy term are given in Table VI.

### E. Hydrogen bond energy ( $E_{\text{HB}}$ )

The water dimer presents a deep minimum ( $\sim 5.0$  kcal/mol) at its global minimum which leads to ice structures having tetrahedral arrangements leading to four strong HBs to each water. The HB involves attraction between a partially positive hydrogen of the donor water molecule to the partially negative lone pair of the acceptor water molecule. Similarly liquid water forms an extensive HB network that is responsible for many of its unique properties. This HB depends strongly on the alignment of the donor hydrogen with respect

to the lone pair of the acceptor molecule. In addition, the lone pairs on each water molecule repel the lone-pair electrons of nearby water molecules. To describe their roles, we include the role of lone pairs on each water molecule explicitly in our FF.

#### 1. Explicit lone pairs

Although, the lone pair electrons do not present distinct directed electron densities, there are minima in the electrostatic potential surfaces around the isolated water molecule that suggest a tetrahedral structure for the water molecule.<sup>55</sup> The accurate Bowman PES for a water dimer also confirms two local minima ( $\sim 109.5^\circ$  apart) during the rotation of donor water around the acceptor water molecule. A tetrahedral bonding network is also found in condensed phases of water such as hexagonal ice. Therefore, we define tetrahedrally arranged lone pairs ( $\text{sp}^3$  hybridized with an angle of  $109.5^\circ$  between them) defined with respect to the two OH bonds.

The explicit lone pairs of the acceptor molecule in a water dimer are shown in Fig. 6. The lone-pair unit vectors ( $\vec{u}_\alpha$  and  $\vec{u}_\beta$ ) are based on the OH bond unit vectors ( $\vec{u}_1$  and  $\vec{u}_2$ ) and the  $\theta_L = 109.5^\circ$  angle between them. First, the  $\vec{u}_1$  and  $\vec{u}_2$  vectors are used to define the perpendicular ( $\vec{u}_p$ ) and bisector ( $\vec{u}_b$ ) unit vectors. The  $\vec{u}_p$  vector is perpendicular to the plane of  $\text{H}_2\text{O}$ ,  $\vec{u}_p = (\vec{u}_1 \times \vec{u}_2)/|\vec{u}_1 \times \vec{u}_2|$ , and  $\vec{u}_b$  is in the plane of  $\text{H}_2\text{O}$  and bisects the HOH angle,  $\vec{u}_b = (\vec{u}_1 + \vec{u}_2)/2$ . Then, the vector that connects acceptor oxygen to the hydrogen of the donor is normalized to give the  $\vec{u}_{\text{OH}}$  unit vector. The  $\alpha$  and  $\beta$  angles (shown in Fig. 6) are defined as the angles that  $\vec{u}_{\text{OH}}$  constructs with  $\vec{u}_\alpha$  and  $\vec{u}_\beta$  unit vectors, respectively. Finally, we use the following three conditions to completely define lone-pair vectors ( $\vec{u}_\alpha$  and  $\vec{u}_\beta$ ):

- the angle between lone-pair vectors and  $\vec{u}_b$  is  $\theta_L/2 = 54.75^\circ$  because  $\vec{u}_b$  bisects the angle between  $\vec{u}_\alpha$  and  $\vec{u}_\beta$ ,
- the angle between lone-pair vectors and  $\vec{u}_p$  is  $\pi/2 - \theta_L/2 = 35.25^\circ$ , and

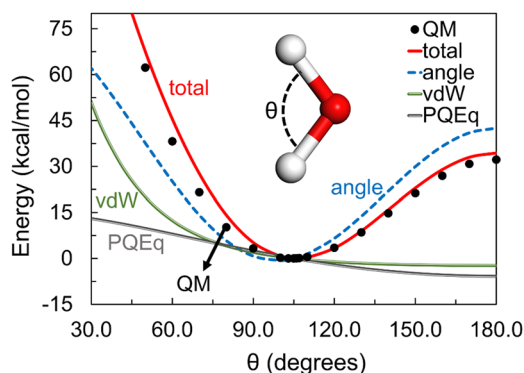


FIG. 5. The energy components and the fit to the HOH angle scan in the water molecule.

TABLE VI. The parameters of the angle energy ( $E_{\text{angle}}$ ) term.

Angle type	$C$ (kcal/mol)	$\theta_0$ (degrees)	$a$	$b$
H–O–H	31.6132	98.8511	3.3562	3.8453

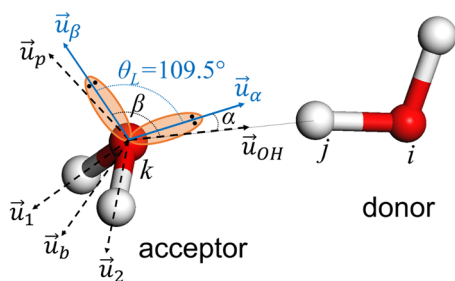


FIG. 6. Schematic defining the explicit lone pairs for the acceptor molecule in the water dimer. The lone-pair vectors ( $\vec{u}_\alpha$  and  $\vec{u}_\beta$ ) are constructed based on the OH bonds vectors ( $\vec{u}_1$  and  $\vec{u}_2$ ) and by solving Eqs. (19)–(21). Note that the angle between the LPs was selected as  $109.5^\circ$  without optimization.

(iii) lone-pair,  $\vec{u}_b$ , and  $\vec{u}_p$  vectors are all in the same plane.

Equations (19)–(21) are written based on these conditions. They are solved simultaneously to obtain  $\vec{u}_\alpha$  and  $\vec{u}_\beta$ ,

$$\vec{u}_b \cdot \vec{u}_\alpha = \cos\left(\frac{\theta_L}{2}\right), \quad (19)$$

$$\vec{u}_p \cdot \vec{u}_\alpha = \cos\left(\frac{\pi - \theta_L}{2}\right), \quad (20)$$

$$\vec{u}_\alpha \cdot (\vec{u}_b \times \vec{u}_p) = 0. \quad (21)$$

The  $E_{HB}$  is defined as the sum of attractive ( $E_{HBatt}$ ) and repulsive ( $E_{HBrep}$ ) terms

$$E_{HB}(r, \theta, \alpha, \beta) = E_{HBatt}(r_{kj}, \alpha, \beta) + E_{HBrep}(r_{ki}, \theta_{ijk}). \quad (22)$$

The  $E_{HBatt}$  term is defined based on lone-pair angles ( $\alpha$  and  $\beta$ ) and the distance ( $r_{kj}$ ) between the acceptor oxygen ( $k$ ) and the hydrogen of the donor molecule ( $j$ ). The  $E_{HBatt}$  is expressed as

$$E_{HBatt}(r_{kj}, \alpha, \beta) = E_{HBatt}(r_{kj})E_{HBatt}(\alpha, \beta), \quad (23)$$

$$E_{HBatt}(r_{kj}) = h_{b1} \left\{ h_{b2} - \left[ e^{-2h_{b3}(r_{kj}-h_{b4})} - h_{b5} \right]^2 \right\}, \quad (24)$$

$$E_{HBatt}(\alpha, \beta) = \frac{h_{b6}}{1 + e^{-h_{b7} \left[ \sin^6\left(\frac{\alpha}{2}\right) + \sin^6\left(\frac{\beta}{2}\right) \right] + h_{b8}}}, \quad (25)$$

where  $h_{b1}$  to  $h_{b8}$  are the adjustable parameters. The  $E_{HBrep}$  term is defined by

$$E_{HBrep}(r_{ik}, \theta_{ijk}) = E_{HBrep}(r_{ik})E_{HBrep}(\theta_{ijk}), \quad (26)$$

$$E_{HBrep}(r_{ik}) = l_{p1} \exp\left(\frac{l_{p2}}{r_{ik}^{l_{p3}}}\right), \quad (27)$$

$$E_{HBrep}(\theta_{ijk}) = 1 + l_{p4} \sin^2\left(\frac{\theta_{ijk}}{2}\right). \quad (28)$$

Here,  $l_{p1}$  to  $l_{p6}$  are the adjustable parameters and  $r_{ik}$  is the oxygen-oxygen distance. We define angle  $\theta_{ijk}$  to be between atoms  $i, j$ , and  $k$  atoms (see Fig. 6) as the average position of the lone pairs is positioned on the center of the oxygen atoms.<sup>55</sup> Since the lone pairs are only defined by vectors, we transfer their force components to the hydrogens of the same molecule via chain derivatives.

For all nonbond energy terms (i.e.,  $E_{electro}$ ,  $E_{vdW}$ , and  $E_{HB}$ ), we use a 7th order taper function to cutoff the forces and energies smoothly at the cutoff distance (matching energy and the first three derivatives at  $r = 0$  and  $r = \text{cutoff}$ ). It is given by

$$T(r_{ij}) = \sum_{\alpha=0}^7 \text{Tap}_\alpha \left( \frac{r_{ij}}{r_{cut}} \right)^\alpha, \quad (29)$$

where  $r_{cut}$  is a cutoff distance and  $\text{Tap}_7 = 20$ ,  $\text{Tap}_6 = -70$ ,  $\text{Tap}_5 = 84$ ,  $\text{Tap}_4 = -35$ ,  $\text{Tap}_3 = 0$ ,  $\text{Tap}_2 = 0$ ,  $\text{Tap}_1 = 0$ , and  $\text{Tap}_0 = 1$ . We used a cutoff distance of  $12.0 \text{ \AA}$  for  $E_{vdW}$  and  $E_{electro}$ . We used a cutoff of  $4.6 \text{ \AA}$  for  $E_{HB}$  to include only the first nearest neighbors in hydrogen bonding.

## 2. HB training set

To obtain  $E_{HB}$  parameters, we started with a minimal training set scanning only four dimer cases. These four cases were selected to establish the range of the parameters and to avoid overfitting. The QM energy for all these cases was obtained using the Bowman PES. The dimer scans are shown in Fig. 7.

- Scan 1: one of the water molecules ( $D_{2h}$  symmetry) is scanned toward the other molecule in the  $xy$  plane. This scan is used to optimize the distance dependent parameters of the  $E_{HBrep}(r_{ik})$  term [Eq. (27)]. Here, it is assumed that there is no angle contribution from  $E_{HBrep}$  [Eq. (28)] and no contribution from  $E_{HBatt}$ .
- Scan 2: the donor water molecule in the most stable water dimer structure is rotated around the  $z$  axis with the oxygen of the donor ( $O_i$ ) as the center of the rotation.
- Scan 3: the donor molecule in the most stable water dimer structure is scanned toward the acceptor with the

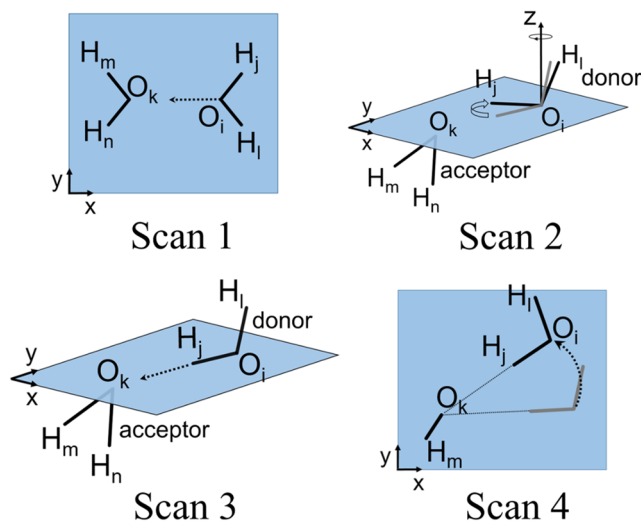


FIG. 7. Four dimer scan cases used in the training set to optimize HB parameters. In Scan 1, the water molecule on the right is scanned toward the water molecule on the left in the  $xy$  plane, in Scan 2, the donor molecule is rotating around the  $z$  axis with the  $O_i$  as the center of rotation, in Scan 3, the donor molecule is scanned toward the acceptor with  $O_iH_jO_k = 180.0^\circ$  and in the  $xy$  plane, and in Scan 4, the donor water molecule is rotating in the  $xy$  plane around  $O_k$  as the center of rotation. In all cases, the bonds and angles are kept fixed during the scans.



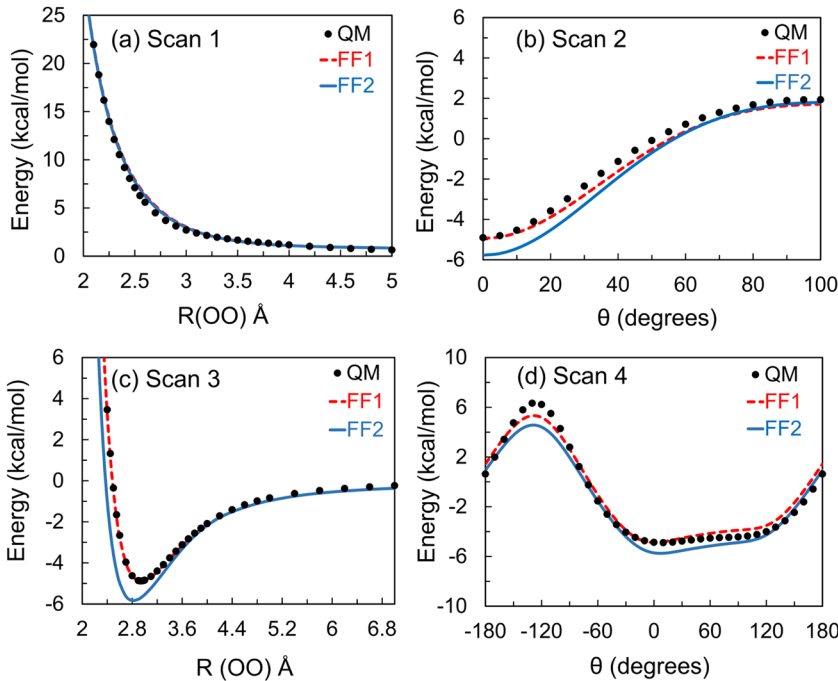


FIG. 8. Energy comparison between RexPoN and QM using the FF1 (red) and FF2 (blue) optimized parameter sets. The FF1 parameter set is obtained using only the 4 scan cases shown in Fig. 7 while for FF2 in addition to 4 scan cases, the 10 low-lying stationary points of water dimers and water clusters,  $(\text{H}_2\text{O})_n$ , containing up to  $n = 19$  clusters were included in the training set.

TABLE VII. The parameters of the  $E_{HBatt}$  energy term [Eqs. (24) and (25)].

Parameters	$h_{b1}$	$h_{b2}$	$h_{b3}$	$h_{b4}$	$h_{b5}$	$h_{b6}$	$h_{b7}$	$h_{b8}$
Value	-22.5000	12.7434	1.1345	0.7861	3.5071	1.1693	3.6922	-0.8921

$\text{O}_i\text{H}_j\text{O}_k$  angle equal to  $180.0^\circ$  during the scan. Scan 2 and Scan 3 are used to obtain the  $E_{HBrep}(\theta_{ijk})$  and  $E_{HBatt}(r_{kj})$  parameters.

- Scan 4: the donor water molecule in the most stable water dimer is rotating in the  $xy$  plane around the oxygen of the acceptor molecule ( $\text{O}_k$ ) as the center of rotation. We used this to optimize the  $E_{HBatt}(\alpha, \beta)$  parameters.

The optimization of the HB parameters for the above four scan cases used conjugate gradients to minimize the root-mean square error (RMSE) of the energy values for each geometry. The energy comparison for the obtained parameter set (FF1) and QM is shown in Fig. 8, which are in very good agreement.

To improve the quality of the force field for different configurations of water dimer interactions, we also included the 10 low-lying stationary point structures (see Fig. S6 of the [supplementary material](#)) of the water dimer in our training set. The reference energies and minimized structures are from the Bowman PES. In addition, to account for possible non-additive HB effects that exist in a bulk system, we expanded our training set by including the global minima of water

clusters,  $(\text{H}_2\text{O})_n$ , containing up to  $n = 19$  waters. The reference energies and global minimized structures were taken from X3LYP<sup>56</sup> hybrid DFT calculations which leads to excellent structures for large water clusters. These calculations were performed using the Jaguar<sup>57</sup> software. The cluster structures are provided in Fig. S7 of the [supplementary material](#). Our optimization of the HB term results in a RMSE difference of 0.65 kcal/mol/water for the above 10 water dimers and 19 water clusters. The comparison between RexPoN and QM for the 4 scan cases after including all dimers and water clusters in the optimization (FF2) is shown in Fig. 8. The energy components of RexPoN for FF2 are shown in Fig. S8 of the [supplementary material](#). The systematic lower energy of the FF2 with respect to FF1 is due the non-additive HB effects that are not captured completely by considering just water dimers. Because of that at the OO distance of 2.76 Å, the water dimer energy computed by FF2 is equal to 5.7 kcal/mol which is close to the experimental ice HB energy<sup>58</sup> of 5.5 kcal/mol (OO = 2.76 Å). The comparison between the computed energies of RexPoN and QM for water dimers and clusters is provided in Table S1 of the [supplementary material](#). The  $E_{HBatt}$  and  $E_{HBrep}$  optimized parameters [Eqs. (23)–(28)] for FF2 are provided in Tables VII and VIII, respectively.

TABLE VIII. The parameters of the  $E_{HBrep}$  energy term [Eqs. (27) and (28)].

Parameters	$l_{p1}$	$l_{p2}$	$l_{p3}$	$l_{p4}$
Value	0.7650	10.2511	2.0214	15.0000

### III. RESULTS

To validate the RexPoN FF, we computed several properties of water and ice including the radial distribution function (RDF), heat of vaporization ( $\Delta H_{vap}$ ), dielectric constant ( $\epsilon$ ),

TABLE IX. Summary of predicted properties from the RexPoN FF compared to experimental data and other popular models.<sup>63–65</sup> The melting temperature  $T_m$  (K) at 1 atm pressure, standard molar entropy  $S^0$  [(J/mol)/K], density  $\rho$  (g/cm<sup>3</sup>), static dielectric constant  $\epsilon$ , heat of vaporization  $\Delta H_v$  (kcal/mol), and self-diffusion coefficient  $D_s$  (cm<sup>2</sup>/s) all at  $T = 298$  K,  $p = 1$  atm. Note that MB-pol is also based only on QM data.

	Expt.	RexPoN	TIP3P	TIP4P-2005	TIP5P	SPC/E	MB-pol
$T_m$	273.15	273.3	146	252	274	215	263.5
$S^0$	69.9	68.43	72.51	57.47	...	60.30	...
$\rho$	0.9965	0.9965	0.98	0.993	0.979	0.994	1.007
$\epsilon$	78.4	76.1	94	58	91	68	68.4
$\Delta H_v$	10.52	10.36	10.05	11.99	10.46	11.79	10.93
$\ln D_s$	-11.24	-10.08	-10.2	-11.27	-11.41	-11.08	-10.46

and diffusivity of water at 298 K, and melting point of ice Ih. A quick comparison for several properties of liquid water and also the melting point of ice Ih between experiment, some of the widely used water models, and RexPoN FF is provided in Table IX.

We integrated the RexPoN model in the LAMMPS<sup>59</sup> MD simulator package and will submit our modifications for inclusion in the official LAMMPS releases. For these water simulations at 298 K, we used a system with 216 molecules per cell, and for ice melting simulations, we used a system with 96 molecules per cell. To control the pressure, we used a barostat with a relaxation time of 1 ps, and to control temperature, we used the Nosé-Hoover thermostat with a damping time of 100 femtosecond (fs).<sup>60,61</sup> We used a rigid model for the water molecules since most practical simulations use rigid models. The OH bond distance and HOH angle are kept fixed at their equilibrium values (OH = 0.9572 Å and HÔH = 104.52°) using SHAKE methodology.<sup>62</sup> We used a time step of 1.0 fs here, but time steps of 2.0 fs could have been used. We used a time step of 1.0 fs to ensure the accuracy for all properties. We expect that using time steps of 2.0 fs would lead to similar accuracy, but we have not yet tested the precision of computed properties using 2.0 fs time steps.

### A. Melting point calculation

To determine the melting point of the ice Ih structure, we started with ice Ih at 0 K and increased temperature gradually to the range of 200–320 K during 10 ps followed by 150 ps constant temperature-volume (MD-NVT) calculations. Then, the cell was relaxed for 120 ps during constant temperature-pressure (MD-NPT) simulations. Then at each  $T$ , the system was further relaxed for another 100 ps (5 cycles each 20 ps) using MD-NVT for performing the two-phase thermodynamics (2PT) analysis.<sup>66,67</sup> The 2PT analysis is used to calculate the thermodynamic properties, including entropy and free energy. 2PT determines the vibrational density of states from the Fourier transform of the velocity autocorrelation function after which corrections are made for the diffusional contributions that lead to a finite density of states at  $\nu = 0$ . 2PT allows the free energy, entropy, and other thermodynamic properties to be determined from short 20 ps MD trajectories. The

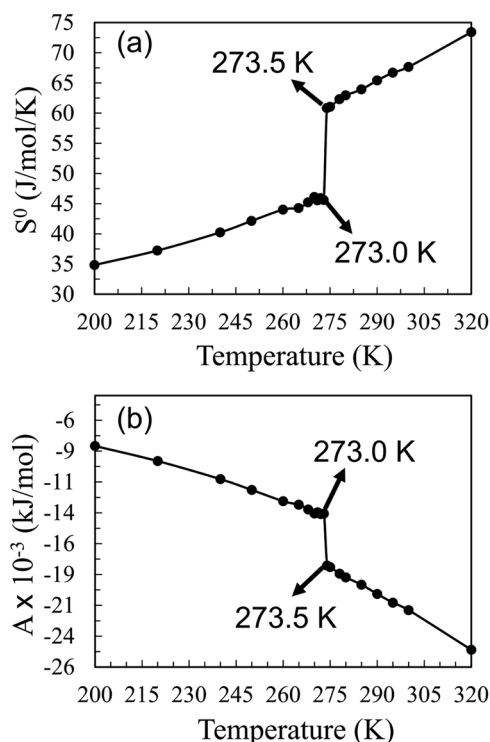


FIG. 9. The (a) absolute entropy and (b) free energy of water as a function of temperature from the 2PT analysis of the MD trajectory. We find a sharp discontinuity between  $T = 273.0$  and  $273.5$  K, corresponding to the melting point of ice Ih.

accuracy of the 2PT method has been validated against expensive *ab initio* methods for the calculation of absolute entropy of liquids.<sup>64</sup> Figure 9 shows the change of the standard molar entropy ( $S^0$ ) and free energy ( $A$ ) of the ice versus temperature. Melting is a first-order transition that results in a discontinuity and sudden change in the slope of the  $S^0$  and  $A$  curves. RexPoN finds that this sharp discontinuity occurs between 273.0 and 273.5 K, indicating a melting point of  $\sim 273.3$  K for ice Ih. This predicted melting point is in excellent agreement with the experimental melting point of 273.15 K at ambient pressure. Also, the free energy curve shows a sharp change at the same temperature.

### B. Liquid density at ambient pressure

We used two methods to evaluate the liquid density of water at 298 K and 1 atm. First, we performed MD simulations in the  $NPT$  ensemble for 1 nanosecond (ns). Since large fluctuations of the pressure during MD- $NPT$  simulations can sometimes skew the  $NPT$  ensemble averages, we also computed the EOS of the liquid water at 298 K for densities in the range of 0.98 to 1.02 gr/cm<sup>3</sup>. For each density, we utilized MD-NVT simulations at 298 K for 1 ns and averaged the pressure values over the last 0.5 ns. The pressure versus density plot is shown in Fig. 10, leading to a density of 0.9965 gr/cm<sup>3</sup> at 1 atm in perfect agreement with the experimental density of 0.9965 gr/cm<sup>3</sup>. Our MD- $NPT$  also resulted in a density of 0.9960 gr/cm<sup>3</sup> also in excellent agreement with experimental density (see Fig. S8 of the [supplementary material](#)).

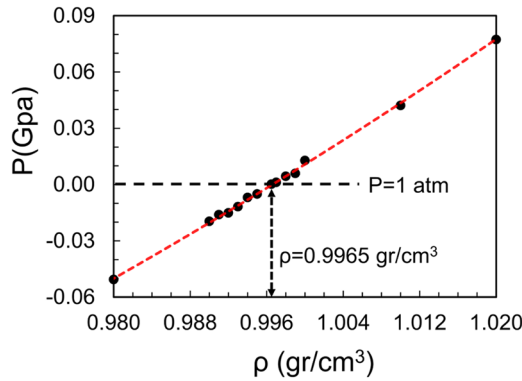


FIG. 10. Equation of state of water predicted at 298 K. The data are fitted to a second order polynomial (dashed line), leading to a density of water at 1 atm of 0.9965 gr/cm<sup>3</sup> at 298 K, in perfect agreement with the experimental density of 0.9965 gr/cm<sup>3</sup>.

### C. Enthalpy of vaporization

We computed the enthalpy of vaporization ( $\Delta H_{\text{vap}}$ ) as the difference between the gas (g) and liquid (l) enthalpies,

$$\Delta H_{\text{vap}}(T) = H(p, T)_g - H(p, T)_l, \quad (30)$$

which can be written in terms of gas and liquid total energies and volumes ( $V_g$  and  $V_l$ ),

$$\Delta H_{\text{vap}}(T) = U_g(T) - U_l(T) + p(V_g - V_l), \quad (31)$$

where  $U_g(T)$  and  $U_l(T)$  are the total energies of the gas and liquid phase, respectively. Equation (31) is further simplified with the assumptions that (i)  $V_l \ll V_g$ , (ii) gas phase is ideal (i.e.,  $pV_g = RT$ ), and (iii) the kinetic energy of water molecules in the liquid and gas phases is equal,

$$\Delta H_{\text{vap}}(T) = E_g(T) - E_l(T) + RT, \quad (32)$$

where  $E_g(T)$  and  $E_l(T)$  are the potential energies of the gas and liquid phase, respectively. It should be noted that Eq. (32) requires no correction for polarization effects since the PQEq model explicitly includes the polarization energy. Using Eq. (32), we computed  $\Delta H_{\text{vap}} = 10.36$  kcal/mol at 298 K compared to be in excellent agreement with experimental  $\Delta H_{\text{vap}} = 10.52$  kcal/mol.

### D. Absolute entropy

We computed the  $S^0$  of liquid at 298 K using 2PT methodology. For this calculation, we extracted the last 100 ps trajectory of the 1 ns MD-NVT simulation described above. The 2PT analysis was performed for five different 20 ps intervals and averaged over the 5 cycles. This leads to  $S^0 = 68.43$  (J/mol)/K, just 2% below that the experimental value of  $S^0 = 69.9$  (J/mol)/K.<sup>68</sup> This accurate value from Rex-PoN for such absolute thermodynamics properties as entropy suggests that this model will reproduce a wide variety of properties of water. The computed  $S^0$  is usually underestimated by other water models. Thus AIMD-PBE (−27%), ReaxFF (−13%), TIP4P-2005 (−18%), TIP4P-ice (−25%), SPC/Fw (−13%), F3C (−12%), effective FF from AIMD (−6%), and TIP4/F (−16%) are compared to experiment.<sup>64</sup>

### E. Static dielectric constant

We computed the static dielectric constant ( $\epsilon$ ) from dipole moment fluctuations

$$\epsilon = \epsilon_\infty + \frac{4\pi}{3k_B TV} (\langle M^2 \rangle - \langle M \rangle^2), \quad (33)$$

where  $k_B$  is the Boltzmann constant,  $T$  is the temperature,  $V$  is the volume,  $M$  is the total dipole moment of the simulation cell, and  $\epsilon_\infty$  is the high frequency or optical dielectric constant. The PQEq formalism allows the direct calculation of  $\epsilon_\infty$  from the classical fluctuation of the electrons (shells) as a measure of the total polarizability of the simulation cell.<sup>17</sup> It can be calculated from the dipole moment fluctuations due to the shell movement around the cores using

$$\epsilon_\infty = 1 + \frac{4\pi}{3k_B TV} (\langle M_s^2 \rangle - \langle M_s \rangle^2), \quad (34)$$

where  $M_s$  indicates the dipole moment resulted from the shell displacement from its core ( $M_s = -\sum_i^N \Delta r_i$ ). To compute the dielectric constant, we performed = 1 ns MD-NVT simulations at 298 K, leading to  $\epsilon_0 = 76.1$  and  $\epsilon_\infty = 1.31$ , in good agreement with the experimental dielectric constant<sup>69</sup> value of 78.4 and the optical dielectric constant<sup>70</sup> of 1.79. The change with time of the  $\epsilon$ ,  $\langle M^2 \rangle$ , and  $\langle M \rangle^2$  are shown in Fig. 11. As shown in this figure, after 300 ps, the values of  $\epsilon$  and  $\langle M^2 \rangle$  reach to a plateau and  $\langle M \rangle^2$  becomes negligible.

### F. Radial distribution function

The molecular distances and orientations in water are mainly determined by the HB network within the first shell of each molecule, making the structural analysis of the first nearest neighbors crucial. We computed the oxygen-oxygen radial distribution function ( $g_{\text{OO}}$ ) and coordination number ( $N_{\text{OO}}$ ) of water at 298 K and ambient pressure over 1 ns MD-NVT simulations. The results together with the experimental data are shown in Fig. 12. The experimental data are based on neutron scattering (Exp1) and combination of X-ray and Neutron experiments (Exp2).<sup>71,72</sup> The first peak of Rex-PoN (2.84 Å) is in excellent agreement with Exp1 (2.85 Å) and Exp2 (2.76 Å). Most empirical water models were adjusted to

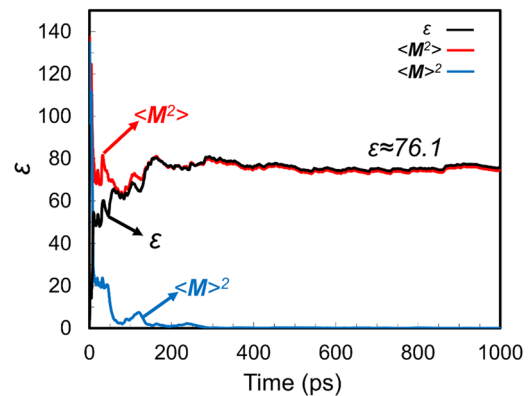


FIG. 11. The change of static dielectric constant ( $\epsilon$ ) and dipole moment fluctuations ( $\langle M^2 \rangle$  and  $\langle M \rangle^2$ ) with time during 1 ns MD-NVT simulations at  $T = 298$  K and ambient pressure. After 300 ps, the values of  $\epsilon$  and  $\langle M^2 \rangle$  reach to a plateau and  $\langle M \rangle^2$  becomes negligible.

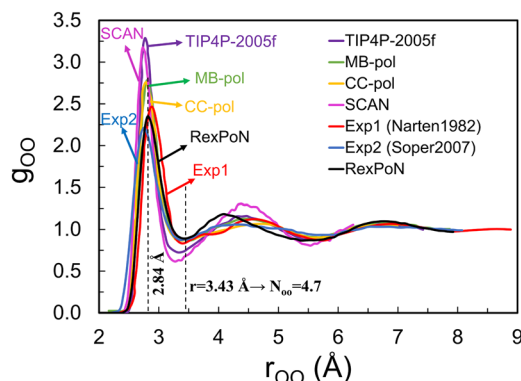


FIG. 12. Comparison of the oxygen-oxygen radial distribution function at  $T = 298$  K and ambient pressure between experiment, RexPoN, TIP4P-2005f,<sup>74</sup> SCAN,<sup>78</sup> CC-pol,<sup>20</sup> and MB-pol<sup>21</sup> water models. The experimental data are based on neutron scattering (Exp1)<sup>71</sup> and combination of X-ray and neutron experiments (Exp2).<sup>72</sup>

have a first peak at  $\sim 2.76$  Å, but they lead to a sharp first peak.<sup>72–74</sup> Figure 12 includes the RDF for the TIP4P/2005 FF<sup>74</sup> (a typical empirical FF) and several *ab initio* based models including SCAN,<sup>75</sup> CC-pol, and MB-pol potentials. All these models lead to a poor fit to the experimental  $g_{OO}$ . Although the *ab initio* models lead to a much better  $g_{OO}$  than the empirical FF. By contrast, RexPoN gives a slope and height of the first peak in excellent agreement with experimental data. Our computed second peak (4.2 Å) is also within the experimental range of 4.2 to 4.7 Å. As shown in Fig. 12, RexPoN leads to a first minimum of the  $g_{OO}$  at 3.43 Å, leading to  $N_{OO} = 4.7$ , in excellent agreement with experiment. This value shows more than four water molecules in the first shell of water, suggesting a broken tetrahedral structure for water.

### G. Self-diffusion coefficient

The self-diffusion coefficient ( $D_s$ ) was derived from the ionic mean-squared displacement (MSD) curve using the 3D diffusion equation of Einstein

$$MSD(t) = \lim_{t \rightarrow \infty} \langle |r_i(t) - r_i(t_0)|^2 \rangle = 6D_s t, \quad (35)$$

where  $r_i(t)$  is the position of particle  $i$  at time  $t$  and  $t_0$  is the time origin and could be any time during the simulation. To provide better sampling, we computed an averaged MSD at time  $t$  with respect to all of the frames smaller than  $t$ . The values of MSD were computed every 1 ps during 1 ns NVT simulations at 298 K and ambient pressure. The change of MSD with time along with its log-log curve is shown in Fig. 13. The log-log curve shows that by 300 ps, the system has reached the Fickian regime with a slope of 1.0, satisfying  $\log(MSD) = \log(t) + \log(6D_s)$ . Our computed  $D_s = 4.21 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$  is higher than the experimental value<sup>76</sup> of  $2.27 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ . From 2PT analysis, we also computed the diffusivity using the density of states at zero frequency,  $DoS(0) = 12 \text{ mN } D_s/(kT)$ , which leads to  $D_s = 4.23 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ . This differs the MSD value by only 0.5%, suggesting that the 2PT analysis can be used to compute the diffusivity of large systems from short simulation times. The larger computed value of diffusivity by RexPoN is puzzling to us. We have done this both for long time scales

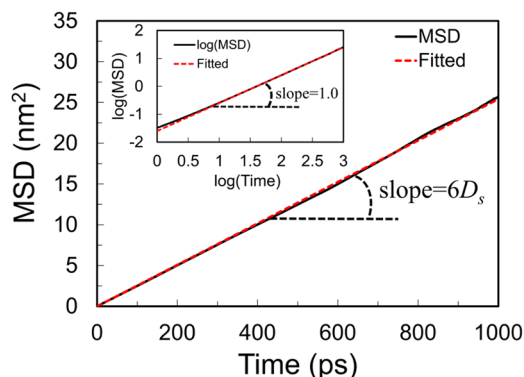


FIG. 13. The change of mean-squared displacement with time during 1 ns MD-NVT simulations at 298 K and 1 atm pressure. The slope of the fitted red line is used to determine the self-diffusivity coefficient ( $D_s$ ) of the liquid according to the Einstein equation. The slope of the log-log curve (inset) becomes equal to 1.0 after about 300 ps.

and from 2PT and obtained the same results. This discrepancy is puzzling because RexPoN leads to good accuracy for the entropy of water [ $S^0 = 68.43 \text{ (J/mol)/K}$ ] compared to the experimental value of  $69.90 \text{ (J/mol)/K}$ . It is known that diffusivity and entropy of a system are directly related to each other. Therefore, one would expect RexPoN to give an accurate value for the diffusivity as it describes the entropy of the system very well.

## IV. CONCLUSIONS

The RexPoN force field based entirely on high quality QM calculations leads to excellent agreement with the experimental properties of the solid and liquid phases of water. This attests to the quality of the QM calculations and to the validity of the RexPoN formalism. We are hopeful that the extension of the RexPoN methodology to other systems, fitting the parameters from QM, may lead to a new generation of FF for very accurate predictions on electrolytes and other materials. We believe that the RexPoN FF for water could provide a significant reduction in the costs for full solvent electrochemical QM calculations by allowing the QM water to be restricted to just a couple of layers.

## SUPPLEMENTARY MATERIAL

The [supplementary material](#) contains the comparison between PQEq and QM charge distributions in the water dimer, the crystal structures of solid hydrogen and solid oxygen, two-body vdW energy curves, energy changes with the volume of the hydrogen crystal, bond energy to the bond order relationship, structures of 10 low-lying stationary points of the water dimer, global minimized structures of 19 water clusters, comparison between RexPoN and QM for the 4 scan cases, comparison between RexPoN and QM energies for water dimers and clusters, the liquid density evaluation using MD-NPT simulations, and the computational cost of RexPoN.

## ACKNOWLEDGMENTS

This work was supported by the Computational Materials Sciences Program funded by the U.S. Department of



Energy, Office of Science, Basic Energy Sciences, under Award No. DE-SC00014607. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation Grant No. ACI-1548562. W.A.G. was supported by NSF (CBET 1512759).

- <sup>1</sup>J. F. Ouyang and R. Bettens, *CHIMIA Int. J. Chem.* **69**(3), 104–111 (2015).
- <sup>2</sup>B. Guillot, *J. Mol. Liq.* **101**(1-3), 219–260 (2002).
- <sup>3</sup>F. Prielmeier, E. Lang, R. Speedy, and H.-D. Lüdemann, *Phys. Rev. Lett.* **59**(10), 1128 (1987).
- <sup>4</sup>P. H. Poole, F. Sciortino, U. Essmann, and H. E. Stanley, *Nature* **360**(6402), 324 (1992).
- <sup>5</sup>C. G. Salzmann, P. G. Radaelli, A. Hallbrucker, E. Mayer, and J. L. Finney, *Science* **311**(5768), 1758–1761 (2006).
- <sup>6</sup>C. G. Salzmann, P. G. Radaelli, E. Mayer, and J. L. Finney, *Phys. Rev. Lett.* **103**(10), 105701 (2009).
- <sup>7</sup>W. L. Jorgensen, *J. Am. Chem. Soc.* **103**(2), 335–340 (1981).
- <sup>8</sup>W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**(2), 926–935 (1983).
- <sup>9</sup>M. W. Mahoney and W. L. Jorgensen, *J. Chem. Phys.* **112**(20), 8910–8922 (2000).
- <sup>10</sup>H. J. Berendsen, J. P. Postma, W. F. van Gunsteren, and J. Hermans, *Intermolecular Forces* (Springer, 1981), pp. 331–342.
- <sup>11</sup>M. W. Mahoney and W. L. Jorgensen, *J. Chem. Phys.* **115**(23), 10758–10768 (2001).
- <sup>12</sup>Y. Wu, H. L. Tepper, and G. A. Voth, *J. Chem. Phys.* **124**(2), 024503 (2006).
- <sup>13</sup>S. W. Rick, S. J. Stuart, and B. J. Berne, *J. Chem. Phys.* **101**(7), 6141–6156 (1994).
- <sup>14</sup>C. Millot and A. J. Stone, *Mol. Phys.* **77**(3), 439–462 (1992).
- <sup>15</sup>E. M. Mas, K. Szalewicz, R. Bukowski, and B. Jeziorski, *J. Chem. Phys.* **107**(11), 4207–4218 (1997).
- <sup>16</sup>C. J. Burnham, J. Li, S. S. Xantheas, and M. Leslie, *J. Chem. Phys.* **110**(9), 4566–4581 (1999).
- <sup>17</sup>G. Lamoureux, A. D. MacKerell, Jr., and B. Roux, *J. Chem. Phys.* **119**(10), 5185–5197 (2003).
- <sup>18</sup>B. T. Thole, *Chem. Phys.* **59**(3), 341–350 (1981).
- <sup>19</sup>X. Huang, B. J. Braams, and J. M. Bowman, *J. Phys. Chem. A* **110**(2), 445–451 (2006).
- <sup>20</sup>R. Bukowski, K. Szalewicz, G. C. Groenenboom, and A. van der Avoird, *J. Chem. Phys.* **128**(9), 094313 (2008).
- <sup>21</sup>V. Babin, C. Leforestier, and F. Paesani, *J. Chem. Theory Comput.* **9**(12), 5395–5403 (2013).
- <sup>22</sup>M. Riera, N. Mardirossian, P. Bajaj, A. W. Götz, and F. Paesani, *J. Chem. Phys.* **147**(16), 161715 (2017).
- <sup>23</sup>M. Sprik, J. Hutter, and M. Parrinello, *J. Chem. Phys.* **105**(3), 1142–1152 (1996).
- <sup>24</sup>C. M. Handley, G. I. Hawe, D. B. Kell, and P. L. Popelier, *Phys. Chem. Chem. Phys.* **11**(30), 6365–6376 (2009).
- <sup>25</sup>A. Shank, Y. Wang, A. Kaledin, B. J. Braams, and J. M. Bowman, *J. Chem. Phys.* **130**(14), 144314 (2009).
- <sup>26</sup>M. Ceriotti, W. Fang, P. G. Kusalik, R. H. McKenzie, A. Michaelides, M. A. Morales, and T. E. Markland, *Chem. Rev.* **116**(13), 7529–7550 (2016).
- <sup>27</sup>S. Habershon, T. E. Markland, and D. E. Manolopoulos, *J. Chem. Phys.* **131**(2), 024501 (2009).
- <sup>28</sup>S. Naserifar, D. J. Brooks, W. A. Goddard III, and V. Cvicek, *J. Chem. Phys.* **146**(12), 124117 (2017).
- <sup>29</sup>J. J. Oppenheim, S. Naserifar, and W. A. Goddard III, *J. Phys. Chem. A* **122**(2), 639–645 (2018).
- <sup>30</sup>S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, *J. Chem. Phys.* **132**(15), 154104 (2010).
- <sup>31</sup>S. Grimme, S. Ehrlich, and L. Goerigk, *J. Comput. Chem.* **32**(7), 1456–1465 (2011).
- <sup>32</sup>J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**(18), 3865 (1996).
- <sup>33</sup>A. D. Becke, *J. Chem. Phys.* **98**(7), 5648–5652 (1993).
- <sup>34</sup>C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**(2), 785 (1988).
- <sup>35</sup>C. Barrett, L. Meyer, and J. Wasserman, *J. Chem. Phys.* **47**(2), 592–597 (1967).
- <sup>36</sup>Y. A. Freiman and H.-J. Jodl, *Phys. Rep.* **401**(1-4), 1–228 (2004).
- <sup>37</sup>P. Loubeyre, R. LeToullec, D. Hausermann, M. Hanfland, R. Hemley, H. Mao, and L. Finger, *Nature* **383**(6602), 702 (1996).
- <sup>38</sup>J. Kohanoff, S. Scandolo, G. L. Chiarotti, and E. Tosatti, *Phys. Rev. Lett.* **78**(14), 2783 (1997).
- <sup>39</sup>M. Städele and R. M. Martin, *Phys. Rev. Lett.* **84**(26), 6070 (2000).
- <sup>40</sup>G. Kresse, *Phys. Rev. B* **54**, 11169 (1996).
- <sup>41</sup>L. F. Lundegaard, G. Weck, M. I. McMahon, S. Desgreniers, and P. Loubeyre, *Nature* **443**(7108), 201 (2006).
- <sup>42</sup>G. Weck, S. Desgreniers, P. Loubeyre, and M. Mezouar, *Phys. Rev. Lett.* **102**(25), 255503 (2009).
- <sup>43</sup>Y. Akahama, H. Kawamura, and O. Shimomura, *Phys. Rev. B* **64**(5), 054105 (2001).
- <sup>44</sup>R. Dovesi, R. Orlando, A. Erba, C. M. Zicovich-Wilson, B. Civalieri, S. Casassa, L. Maschio, M. Ferrabone, M. De La Pierre, and P. D’Arco, *Int. J. Quantum Chem.* **114**(19), 1287–1317 (2014).
- <sup>45</sup>J. Heyd, J. E. Peralta, G. E. Scuseria, and R. L. Martin, *J. Chem. Phys.* **123**(17), 174101 (2005).
- <sup>46</sup>H. Kim, J.-M. Choi, and W. A. Goddard III, *J. Phys. Chem. Lett.* **3**(3), 360–363 (2012).
- <sup>47</sup>Y. Liu and W. A. I. Goddard, *Mater. Trans.* **50**(7), 1664–1670 (2009).
- <sup>48</sup>L. Liu, Y. Liu, S. V. Zybin, H. Sun, and W. A. Goddard III, *J. Phys. Chem. A* **115**(40), 11016–11022 (2011).
- <sup>49</sup>A. K. Rappé, C. J. Casewit, K. Colwell, W. Goddard III, and W. Skiff, *J. Am. Chem. Soc.* **114**(25), 10024–10035 (1992).
- <sup>50</sup>F. Birch, *Phys. Rev.* **71**(11), 809 (1947).
- <sup>51</sup>A. C. Van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, *J. Phys. Chem. A* **105**(41), 9396–9409 (2001).
- <sup>52</sup>J. S. Sims and S. A. Hagstrom, *J. Chem. Phys.* **124**(9), 094101 (2006).
- <sup>53</sup>L. Bytautas and K. Ruedenberg, *J. Chem. Phys.* **132**(7), 074109 (2010).
- <sup>54</sup>F. R. Gilmore, *J. Quant. Spectrosc. Radiat. Transfer* **5**(2), 369–390 (1965).
- <sup>55</sup>A. Kumar, S. R. Gadre, N. Mohan, and C. H. Suresh, *J. Phys. Chem. A* **118**(2), 526–532 (2014).
- <sup>56</sup>J. T. Su, X. Xu, and W. A. Goddard, *J. Phys. Chem. A* **108**(47), 10518–10526 (2004).
- <sup>57</sup>Schrödinger Release 2015-2: Jaguar, version 8.8, Schrödinger, LLC, New York, NY, 2015.
- <sup>58</sup>F. H. Stillinger, *Science* **209**(4455), 451–457 (1980).
- <sup>59</sup>S. Plimpton, *J. Comput. Phys.* **117**(1), 1–19 (1995).
- <sup>60</sup>W. G. Hoover, *Phys. Rev. A* **31**(3), 1695 (1985).
- <sup>61</sup>S. Nosé, *J. Chem. Phys.* **81**(1), 511–519 (1984).
- <sup>62</sup>J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, *J. Comput. Phys.* **23**(3), 327–341 (1977).
- <sup>63</sup>S. K. Reddy, S. C. Straight, P. Bajaj, C. Huy Pham, M. Riera, D. R. Moberg, M. A. Morales, C. Knight, A. W. Götz, and F. Paesani, *J. Chem. Phys.* **145**(19), 194504 (2016).
- <sup>64</sup>T. A. Pascal, D. Schärff, Y. Jung, and T. D. Kühne, *J. Chem. Phys.* **137**(24), 244507 (2012).
- <sup>65</sup>C. Vega and J. L. Abascal, *Phys. Chem. Chem. Phys.* **13**(44), 19663–19688 (2011).
- <sup>66</sup>S.-T. Lin, P. K. Maiti, and W. A. Goddard III, *J. Phys. Chem. B* **114**(24), 8191–8198 (2010).
- <sup>67</sup>T. A. Pascal, S.-T. Lin, and W. A. Goddard III, *Phys. Chem. Chem. Phys.* **13**(1), 169–181 (2011).
- <sup>68</sup>J. Cox, D. Wagman, and V. Medvedev, USA Google Scholar, New York, 1989.
- <sup>69</sup>D. P. Fernández, Y. Mulev, A. Goodwin, and J. L. Sengers, *J. Phys. Chem. Ref. Data* **24**(1), 33–70 (1995).
- <sup>70</sup>A. Buckingham, *Proc. R. Soc. A* **238**(1213), 235–244 (1956).
- <sup>71</sup>A. Narten, W. Thiessen, and L. Blum, *Science* **217**(4564), 1033–1034 (1982).
- <sup>72</sup>A. Soper, *J. Phys.: Condens. Matter* **19**(33), 335206 (2007).
- <sup>73</sup>H. Berendsen, J. Grigera, and T. Straatsma, *J. Phys. Chem.* **91**(24), 6269–6271 (1987).
- <sup>74</sup>M. A. González and J. L. Abascal, *J. Chem. Phys.* **135**(22), 224516 (2011).
- <sup>75</sup>M. Chen, H.-Y. Ko, R. C. Remsing, M. F. C. Andrade, B. Santra, Z. Sun, A. Selloni, R. Car, M. L. Klein, and J. P. Perdew, *Proc. Natl. Acad. Sci. U. S. A.* **114**(41), 10846–10851 (2017).
- <sup>76</sup>R. Mills, *J. Phys. Chem.* **77**(5), 685–688 (1973).
- <sup>77</sup>C. McBride, C. Vega, E. G. Noya, R. Ramírez, and L. M. Sesé, *J. Chem. Phys.* **131**(2), 024506 (2009).
- <sup>78</sup>J. Wiktor, F. Ambrosio, and A. Pasquarello, *J. Chem. Phys.* **147**, 216101 (2017).